THE BENEFITS OF USING PSEUDONYMISED DATA

20 November 2014

Uwe W Fiedler,
Chief Privacy Officer & VP DP

DISCLAIMER: The views expressed in the workshop slides are purely those of the presenter and may not in any circumstances be regarded as providing legal advice or stating an official position of PAREXEL International Corporation.
AGENDA

• What are pseudonymized data?

• Pseudonymized data in Medical Research
  • Ethical Background
  • Data Flow
  • What are pseudonymized clinical research data?

• Benefit of pseudonymized data in daily life

• Summary
WHAT ARE PSEUDONYMIZED DATA?
§3 (6a) "Aliasing (pseudonymization / key-coding)" means replacing a person's name and other identifying characteristics with a label, in order to preclude identification of the data subject or to render such identification substantially difficult.

Question – do we have here pseudonymized data?

The first post-war Chancellor of Germany was born in Cologne.
WP 136 - Opinion 4/2007 on the concept of personal data:

“… Pseudonymisation is the process of disguising identities. The aim of such a process is to be able to collect additional data relating to the same individual without having to know his identity. …

Pseudonymisation can be done in a retraceable way by using correspondence lists for identities and their pseudonyms or by using two-way cryptography algorithms for pseudonymisation. …

Retraceably pseudonymised data may be considered as information on individuals which are indirectly identifiable. Indeed, using a pseudonym means that it is possible to backtrack to the individual, so that the individual’s identity can be discovered, but then only under predefined circumstances. In that case, although data protection rules apply, the risks at stake for the individuals with regard to the processing of such indirectly identifiable information will most often be low, so that the application of these rules will justifiably be more flexible than if information on directly identifiable individuals were processed…. …”
WP 203 - Opinion 03/2013 on purpose limitation:

“… Once again, it is helpful to distinguish different scenarios for further analysis:

» Scenario 1: unidentifiable personal data: data are anonymised or aggregated in such a way that there is no remaining possibility to (reasonably) identify the data subjects.

» Scenario 2: indirectly identifiable personal data: lower level of aggregation, partial anonymization, pseudonymisation or key-coded data.

» Scenario 3: situations where directly identifiable personal data are needed due to the nature of the research.

_85_ Article 2(a) of the Directive defines 'personal data' as 'any information relating to an identified or identifiable natural person ("data subject"); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, psychological, mental, economic, cultural or social identity'. See also Opinion 4/2007 on the concept of personal data, adopted on 20.06.2007 (WP 136), especially on p. 12-21 (discussing 'pseudonymised data', 'key-coded data' and 'anonymous data' on p. 18-21). The issue of information 'relating to' an individual is discussed on p. 9-12.

• Introduce the concept that in regard to personal data it is possible to differentiate between the identifiable and content part of a data set:
  – **payload data** is the part of the Personal Data that contain characteristics that do not allow unique identification of the data subject. Therefore the payload data themselves would only contain anonymous data
  – **identifying data** is the part of the Personal Data that contain a set of characteristics that allow unique identification of the data subject

• Define, therefore, pseudonymization as a particular type of anonymization that both removes the association with a data subject and adds an association between a particular set of characteristics relating to the data subject and one or more pseudonyms

• Note that the conceptual distinction between “identifying data” and “payload data” can lead to contradictions. This is the case when, for example, directly identifying data are considered “payload data”
• The European Commission introduced in their proposal for an EU Data Protection Regulation only indirectly pseudonymized data in Article 83:

\[(b) \text{ data enabling the attribution of information to an identified or identifiable data subject is kept separately from the other information as long as these purposes can be fulfilled in this manner.}\]

• The European Parliament legislative resolution of 12 March 2014 define pseudonymized data as follows:

\[\text{Article 4 (2a) 'pseudonymous data' means personal data that cannot be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution;}\]

• Unfortunately, the proposal of the European Parliament did not introduce a clear framework for the use of pseudonymized data
• The European Council introduced in their proposal for a EU Data Protection Regulation so far, at least a small framework for the use of pseudonymized data:

  • Article 23 recognizes pseudonymization as data protection by design and by default
  • Article 30 recognizes pseudonymization as appropriate technical and organizational measures
  • Article 32 exempts pseudonymized data from security breach notification to data subject
  • Article 38 emphasizes the implementation of Code of Conducts for the use of pseudonymized data
  • During the Greek presidency, the Council updated Article 26 to also allow Processors to demonstrate sufficient guarantees for their implementation of appropriate technical and organizational measures by adherence of the Processor to codes of conducts

• These ideas are underlined by their recently published “Handbook on European Data Protection Law”
2.1. Personal data

» …

» In contrast to anonymised data, pseudonymised data are personal data.

2.1.3. Anonymised and pseudonymised data

… As pseudonymisation of data is one of the most important means of achieving data protection on a large scale, where it is not possible to entirely refrain from using personal data, the logic and the effect of such action must be explained in more detail.

…

Personal data with encrypted identifiers are used in many contexts as a means to keep secret the identity of persons. This is particularly useful where data controllers need to ensure that they are dealing with the same data subjects but do not require, or ought not to have, the data subjects’ real identities. This is the case, for example, where a researcher studies the course of a disease with patients, whose identity is known only to the hospital where they are treated and from which the researcher obtains the pseudonymised case histories. Pseudonymisation is therefore a strong link in the armoury of privacy-enhancing technology. It can function as an important element when implementing privacy by design. This means having data protection built into the fabric of advanced data-processing systems.

• The discussion of the past years clarified that pseudonymized data are “a form of“ personal data

• Unfortunately it is still unclear which content could transform anonymized data into pseudonymized data and vice versa

• The European Commission recently published their Handbook on European Data Protection Law that strongly suggests the increasing use of pseudonymized data

• Besides the hope on an improved Data Protection Regulation, another potential solution for this problem could therefore be the implementation of sector specific Code of Conducts that could harmonize the sector related privacy requirements within the EU (this was already possible under the 1995 Directive and still seems to be possible under a possible Regulation)
PSEUDONYMIZED DATA IN MEDICAL RESEARCH
ETHICAL BACKGROUND

Declaration of Helsinki
as a statement of ethical principles for medical research involving human subjects

Clinical Development
New and better drugs for patients

Data Protection & Privacy
Protection of personal data as human right

Pharmacovigilance
Safety of the Patients
ETHICAL AND REGULATORY REQUIREMENT

• 1964 - World Medical Association (WMA) developed the Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects¹

  9. It is the duty of physicians who are involved in medical research to protect the life, health, dignity, integrity, right to self-determination, privacy, and confidentiality of personal information of research subjects. The responsibility for the protection of research subjects must always rest with the physician or other health care professionals and never with the research subjects, even though they have given consent.

  24. Every precaution must be taken to protect the privacy of research subjects and the confidentiality of their personal information.

• Medical research contains the risk of adverse reactions – Therefore, the research subject must be re-identifiable at each processing step by protecting the confidentiality of their personal data

• This is only possible via the use of pseudonymized data

¹ 64th WMA General Assembly, Fortaleza, Brazil, October 2013 - http://www.wma.net/en/30publications/10policies/b3/
WHAT ARE PSEUDONYMIZED CLINICAL RESEARCH DATA?

• The German Federal Drug Law (AMG) require the use of pseudonymized data:

§40 (2a) …The person concerned shall be informed especially of the fact that:

1. where necessary, the recorded data:

   a) will be kept available for inspection by the supervisory authority or the sponsor's representative in order to verify the proper conduct of the clinical trial (means access to directly identifiable sensitive personal data),

   b) will be passed on in a pseudonymized version to the sponsor or to an agency commissioned by the latter for the purpose of scientific evaluation,

• The German Federal Data Protection Act define pseudonymized data as follows:

§3 (6a) "Aliasing (pseudonymization / key-coding)" means replacing a person's name and other identifying characteristics with a label, in order to preclude identification of the data subject or to render such identification substantially difficult.
WHAT ARE PSEUDONYMIZED CLINICAL RESEARCH DATA?

• As an example, the Berlin Ethic Commission recognizes research subject data only as pseudonymized if the data contains:
  • no initials (neither of the first name, nor of the surname) and
  • if necessary month and year (01.2001) but no full Date of Birth (like 01.01.2001), in order to permit age verification.

• The actual Data Protection Directive does not define pseudonymized data but in 2007 the Article 29 WP stated in their WP 136

  "…The pharmaceutical company has construed the means for the processing, included the organizational measures and its relations with the researcher who holds the key in such a way that the identification of individuals is not only something that may happen, but rather as something that must happen under certain circumstances. The identification of patients is thus embedded in the purposes and the means of the processing.

  …
In this case, one can conclude that such key-coded data constitutes information relating to identifiable natural persons for all parties that might be involved in the possible identification and should be subject to the rules of data protection legislation. …"
WHAT ARE PSEUDONYMIZED CLINICAL RESEARCH DATA?

• Let´s have a look to the other side of the Atlantic

• Since 1996 the US Federal Health Insurance Portability and Accountability Act of 1996 (HIPAA) differentiate between:
  • Protected Health Information (Patient name + medical records = sensitive personal data)
  • Limited Data Set (a form of pseudonymized data still covered by HIPAA)
  • De-identified Data (a form of anonymized data not any longer covered by HIPAA)

• Even if the majority of medical research activities of pharmaceutical companies are not covered by HIPAA, this classification framework is also used there as de-facto standard for pseudonymization of data

• The reason for this is a clear definition of data elements that must be stripped of Protected Health Information to receive a Limited Data Set
WHAT ARE PSEUDONYMIZED CLINICAL RESEARCH DATA?

• A Limited Data Sets (LDS) lacks 16 of the 18 identifiers itemized by the HIPAA Privacy Rule. Specifically, a LDS does NOT include the following identifiers:

<table>
<thead>
<tr>
<th>Name</th>
<th>Postal address information, other than town or city, State, and zip codes;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Telephone numbers;</td>
<td>Fax numbers;</td>
</tr>
<tr>
<td>Electronic mail addresses;</td>
<td>Social Security Numbers;</td>
</tr>
<tr>
<td>Medical record numbers;</td>
<td>Health plan beneficiary numbers;</td>
</tr>
<tr>
<td>Account numbers;</td>
<td>Certificate/license numbers;</td>
</tr>
<tr>
<td>Vehicle identifiers and serial numbers,</td>
<td>Device identifiers and serial numbers;</td>
</tr>
<tr>
<td>including license plate numbers;</td>
<td></td>
</tr>
<tr>
<td>Web Universal Resource Locators (URLs);</td>
<td>Internet Protocol (IP) address numbers;</td>
</tr>
<tr>
<td>Biometric identifiers, including finger and voice prints; and</td>
<td>Full face photographic images and any comparable images.</td>
</tr>
</tbody>
</table>

• An LDS may contain, for example:

<table>
<thead>
<tr>
<th>Dates of birth</th>
<th>Dates of death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dates of service</td>
<td>Town or city</td>
</tr>
<tr>
<td>State</td>
<td>Zip code</td>
</tr>
</tbody>
</table>
• So far the definition of pseudonymized data is unclear in regard to medical research activities within the EU

• It is also unclear to what extent pseudonymized data sets could contain more identifiable data elements like initials and/or full DOB - e.g. for safety reasons to solve pregnant men problem

• The pragmatically approach taken by HIPAA with the introduction of a Limited Data Set that is still protected by HIPAA may be a solution

• But why not implement a pharmaceutical Code of Conduct that would also define the content and purpose for the use of key-coded research subject data?
BENEFIT OF PSEUDONYMIZED DATA IN DAILY LIFE
PSEUDONYMIZATION = PRIVACY BY DESIGN?

• Use of Employee IDs instead of employee names in surveys
  • Benefits:
    – third party service provider do not receive directly identifiable personal data as long as service provider cannot link Employee IDs back to employee name
  • Problems:
    – Majority of companies still use First_Name.Last_Name@Company as Email address so that service provider would receive employee names + could link them to the Employee IDs
• Suggestion:
  – Companies should use pseudonymized email addresses instead of names
PSEUDONYMIZATION = PRIVACY BY DESIGN?

- Use of Trusted Third Parties for pseudonymization
- Quality Assurance in Kidney Disease (QUASINiere) – Germany 1995
Secure Pseudonym Storage and Keeping of Blood and Tissue Samples intended for genetic analyses (double-pseudonymization of blood samples GENOMatch) – Germany 2003

Check in Patient ID + BC1 and then remove Patient ID from sample

Register BC1 / BC2 assignment and then replace BC1 with BC2

Sample management solely with BC2
PSEUDONYMIZATION = PRIVACY BY DESIGN?

• Received privacy certificates from DPA Schleswig-Holstein (https://www.datenschutzzentrum.de/audit/kurzgutachten/a0303/a0303_engl.htm)

… The double pseudonym process that is envisioned in the concept guarantees the implementation of the requirements of the §§ 3a, 9 BDSG and the above mentioned guidelines. … With pseudonyms, it can be avoided that authorities acquire the identity of the trial subject in the context of pharmaceutical inspections. …

The concept also ensures in isolated cases when informed consent is withdrawn after a successful pseudonym process, that a deletion of the data or a destruction of the samples can take place and that any requests for information from the concerned parties or any notification of the concerned can still be satisfied. At the same time it can also be guaranteed that the trial subject remains anonymous to those participating in the research project during the complete scientific processing of the samples and the data by way of the secure double pseudonym procedure. Even in the case that genetic or clinical data fall into the hands of an unauthorized party, the activation of the matching procedure requires two separate and independent sites. …
CONCLUSION
PSEUDONYMIZATION = PRIVACY BY DESIGN!

• The use of pseudonymized and, therefore, indirect identifiable personal data is a long standing Privacy by Design approach to reduce the risk of identifiability (see also http://www.privacybydesign.ca/)

• If the recipient is unable (either by technology or by binding contracts) to re-identify the data subject, this could dramatically reduce the legal requirements for an international data transfer

• Missing piece is the acknowledgment of this Privacy by design approach in laws

• The attempt of the European Council to make the use of pseudonymized data more attractive for controller, as well as, for processor should be understood by them as a signal to also invest into this Privacy by Design model
THANK YOU